












The co-existence of cold activity and thermal stability in an Antarctic GH42 β -galactosidase relies on its hexameric quaternary arrangement

Marco Mangiagalli¹ , Michela Lapi² , Serena Maione¹ , Marco Orlando¹ , Stefania Brocca¹ , Alessandra Pesce³ , Alberto Barbiroli⁴ , Carlo Camilloni² , Sandra Pucciarelli⁵ , Marina Lotti¹  and Marco Nardini² 

¹ Department of Biotechnology and Biosciences, University of Milano-Bicocca, Italy

² Department of Biosciences, University of Milano, Italy

³ Department of Physics, University of Genova, Italy

⁴ Department of Food, Environmental and Nutritional Sciences, University of Milano, Italy

⁵ School of Biosciences and Veterinary Medicine, University of Camerino, Italy

Keywords

cold adaptation; cooperativity; enzyme kinetics; glycoside hydrolase; psychrophilic enzyme

Correspondence

M. Lotti, Department of Biotechnology and Biosciences, University of Milano-Bicocca, P.zza della Scienza 2, 20126 Milano, Italy
 Tel: +39 02 6448 3527

E-mail: marina.lotti@unimib.it

M. Nardini, Department of Biosciences, University of Milano, Via Celoria 26, 20133 Milano, Italy

Tel: +39 02 503 14893

E-mail: marco.nardini@unimi.it

(Received 3 March 2020, revised 25 April 2020, accepted 29 April 2020)

doi:10.1111/febs.15354

To survive in cold environments, psychrophilic organisms produce enzymes endowed with high specific activity at low temperature. The structure of these enzymes is usually flexible and mostly thermolabile. In this work, we investigate the structural basis of cold adaptation of a GH42 β -galactosidase from the psychrophilic *Marinomonas* ef1. This enzyme couples cold activity with astonishing robustness for a psychrophilic protein, for it retains 23% of its highest activity at 5 °C and it is stable for several days at 37 °C and even 50 °C. Phylogenetic analyses indicate a close relationship with thermophilic β -galactosidases, suggesting that the present-day enzyme evolved from a thermostable scaffold modeled by environmental selective pressure. The crystallographic structure reveals the overall similarity with GH42 enzymes, along with a hexameric arrangement (dimer of trimers) not found in psychrophilic, mesophilic, and thermophilic homologues. In the quaternary structure, protomers form a large central cavity, whose accessibility to the substrate is promoted by the dynamic behavior of surface loops, even at low temperature. A peculiar cooperative behavior of the enzyme is likely related to the increase of the internal cavity permeability triggered by heating. Overall, our results highlight a novel strategy of enzyme cold adaptation, based on the oligomerization state of the enzyme, which effectively challenges the paradigm of cold activity coupled with intrinsic thermolability.

Database

Structural data are available in the Protein Data Bank database under the accession number [6Y2K](https://doi.org/10.22424/6Y2K).

Abbreviations

CAZy, carbohydrate active enzyme (database); CD, circular dichroism (spectroscopy); GH, glycoside hydrolase; K_{SV} , Stern–Volmer constant; MD, molecular dynamic; M- β Gal, β -galactosidase from an Antarctic bacterium of *Marimononas* ef1; ONPG, *ortho*-nitrophenyl- β -galactoside; PB, phosphate buffer; R_g , radius of gyration; R_h , hydrodynamic radius; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuations; SASA, solvent-accessible surface area; SEC-LS, size-exclusion chromatography combined with light scattering detection; T_{opt} , optimum temperature of the catalysis.

Introduction

Organisms endemic to environments extreme as for temperature, pressure, pH, and other conditions (so-called extremophiles) developed adaptive strategies suited to their survival. Life in these harsh conditions relies on changes in the cell components, the study of which is relevant for understanding the mechanisms of environment-driven evolution at the molecular level. Besides, extremophiles are valuable sources of molecules and compounds exploitable in several fields of application [1–4].

Among extremophiles, Antarctic marine microorganisms are permanently exposed to temperatures close to the freezing point of water and evolved enzymes active at low temperature, referred to as cold-active enzymes [5,6]. Several such enzymes are described in literature [7], and their ability to cope with the overall impairment of chemical reactions occurring at cold temperature has been ascribed to the peculiar conformational flexibility of local regions (e.g., the active site or protein surface areas). Increased flexibility relies on structural adaptations including peculiar amino acid composition, weakening of molecular interactions (e.g., hydrogen bond, salt bridges), loops extension, and lower metal binding affinity. Flexibility translates into specific biochemical properties such as low activation enthalpy, low affinity for substrates, low optimal temperature of catalysis (T_{opt} , 20–30 °C) and high specific activity at low temperature [7–10]. A frequent side effect of flexibility is the intrinsic thermolability of several psychrophilic enzymes, the majority of which undergo inactivation and denaturation at milder temperature, compared with that of their mesophilic and thermophilic homologues [7–9,11]. Such a trade-off between flexibility and stability is often interpreted as the necessary price to pay for cold activity. This paradigm, however, is questioned by a few recent reports [12–15].

In this work, we investigated the properties of a cold-active glycoside hydrolase (GH). GHs are highly diversified in specificity, structures, mechanisms and functions, which reflects in their complex classification. At present, the database of carbohydrate-active enzymes (CAZy) includes 166 GH families [16–18]. β -Galactosidases (EC 3.2.1.23) catalyze the hydrolysis of β -glycosidic bonds of β -galactosides to give galactose. They are widespread enzymes and have been identified in several organisms, including animals, plants, fungi, yeasts, bacteria, and Archaea [19]. In the CAZy database, β -galactosidases are classified into four classes, GH1, GH2, GH35, and GH42, according to their

sequence similarity [17]. Although all β -galactosidases share the catalytic mechanism and the catalytic domain shaped as an $(\alpha/\beta)_8$ (TIM) barrel [20–25], they differ in substrate specificity. GH1 and GH2 β -galactosidases are specific for lactose, whereas GH35 and GH42 enzymes are specific toward β -1-3, β -1-4, and β -1-6 galactosidic bonds [26]. In GH42 enzymes, the catalytic center consists of two residues of glutamic acid, which act through a retaining mechanism [20,27–30]. Besides the catalytic domain, the overall architecture shows other accessory domains, such as the so-called trimerization domain required for oligomerization. GH42 β -galactosidases of known 3D structure are organized in trimeric quaternary structures [20,27–30]. The GH42 family includes enzymes from Eubacteria, Archaea, and Eukaryota isolated from extremophiles, including psychrophilic [14,31–36] and thermophilic microorganisms [30,37,38], which are particularly useful to investigate about molecular strategies of thermal adaptation.

GH42s have been identified in several psychrophiles, including *Halobacterium lacusprofundi* [32], *Arthrobacter* spp. [33,36], *Planococcus* spp. [34,35], *Rahnella* sp. R3 [31], and *Marinomonas* sp. BSI20414 [14]. These enzymes are active at low temperatures and exhibit a large diversity in their T_{opt} (30–60 °C) [14,31–36], thus suggesting that present psychrophilic GH42s might have originated following different pathways of evolution. Unfortunately, the paucity of high-resolution structural data does not allow shedding light on the molecular basis of cold adaptation in this group of glycosidases. To date, the only solved structure of a GH42 cold-active β -galactosidase is that of *Rahnella* sp. R3. This enzyme is a highly flexible homotrimer that displays a T_{opt} of 35 °C and undergoes complete inactivation within 5 min at 55 °C [28,31].

Here, we report the biochemical and structural characterization of a β -galactosidase from *Marinomonas* ef1 (M- β Gal), a species isolated in Antarctica. This enzyme is atypical as it couples activity at 5 °C with unusual thermal stability. Moreover, it shows allosteric behavior at high temperature. To investigate in depth the molecular basis of M- β Gal features, we combined a set of complementary techniques including biochemical and biophysical analysis, X-ray crystallography, and molecular dynamics (MD) simulations. Our results show that M- β Gal is a novel member of the GH42 family organized in a hexameric structure obtained by the juxtaposition of two trimers, and suggest a pivotal role of the quaternary arrangement for the peculiar behavior of the enzyme.

Results

M- β Gal combines cold activity with high optimal temperature and thermostability

The M- β Gal coding sequence (accession number: WP_100635792) was identified in the genome of *Marinomonas* e1, isolated from a microbial consortium previously sampled in Antarctica [39]. Sequence analysis indicates a putative protein length of 656 amino acids. According to InterPro [40] and Pfam [41], the enzyme belongs to the GH42 glycosidase family. The recombinant protein was produced in *Escherichia coli* cells and purified by affinity chromatography with a final yield of 15–20 mg of pure protein per liter of culture.

Activity assays using *ortho*-nitrophenyl- β -galactoside (ONPG) as the substrate highlighted that M- β Gal has an optimal pH of 6.0 and retains 80% of its activity in

the pH range 5.0–7.0 (Fig. 1A). M- β Gal is salt-tolerant and at 2.4 M NaCl still displays 60% of its highest activity (Fig. 1B). This property might reflect the physiological adaptation of *Marinomonas* e1 to its natural environment (Ross Sea, salinity of $35 \text{ g}\cdot\text{kg}^{-1} \approx 0.6 \text{ M NaCl}$) [42]. The highest catalytic activity was recorded at 55 °C (Fig. 1C), while at 5 °C the relative activity approached 23% of the maximum ($979 \pm 41 \text{ U}\cdot\text{mg}^{-1}$). The unfolding transition midpoint (T_m) was determined by circular dichroism (CD) analysis at a fixed wavelength (205 nm) and temperature in the range from 5 to 90 °C (Fig. 1D). In this experiment, M- β Gal revealed surprisingly high thermal stability with a T_m of 68.7 ± 1.4 °C. The onset of kinetic inactivation (Fig. 1C) was observed at a temperature ~ 13 °C lower than the unfolding temperature. Uncoupling surmises that the active site is more labile than the overall structure, as it was reported for several other cold-active enzymes [7,8,43].

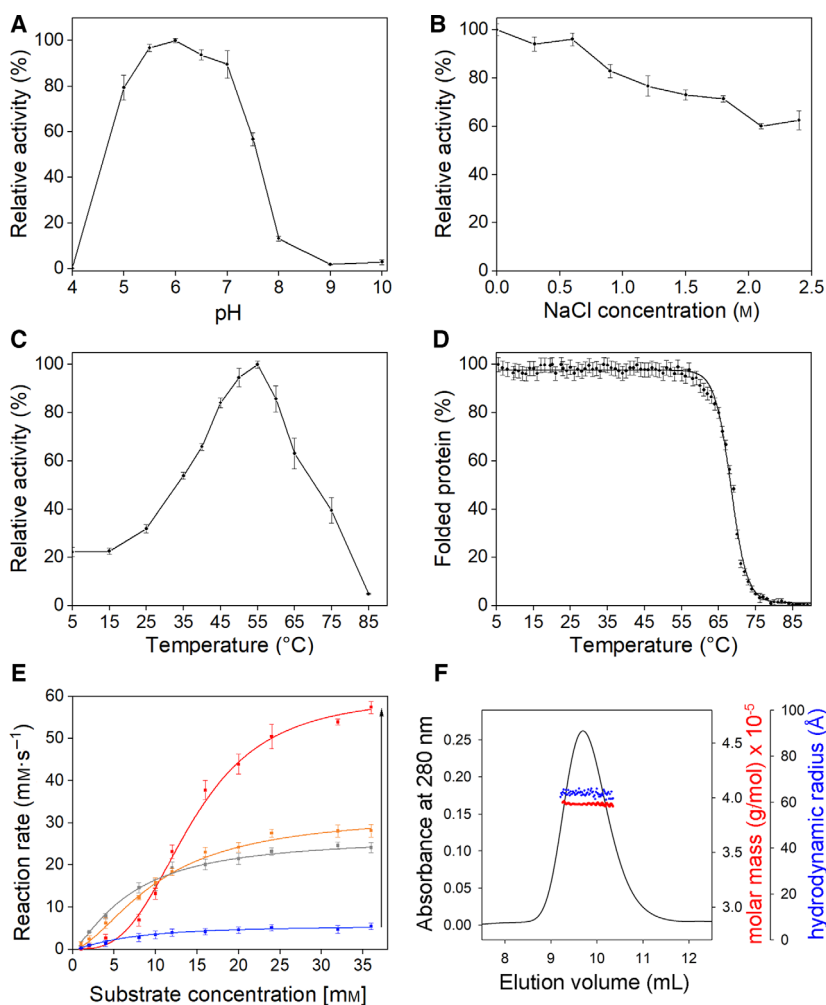


Fig. 1. Biochemical and biophysical characterization of M- β Gal. The effects pH (A), salt (B), and temperature (C) on the activity of M- β Gal were detected using ONPG as a substrate. (D) Thermal stability of M- β Gal. Ellipticity values were recorded at 205 nm during heating from 25 to 90 °C. The initial CD signal was taken as 100% for normalization. (E) The steady-state kinetics of M- β Gal were measured at 5 °C (blue), 25 °C (gray), 40 °C (orange), and 55 °C (red) at different concentrations of ONPG. Error bars indicate standard deviations on three independent experiments (A–E). (F) Oligomerization state of M- β Gal analyzed by SEC (absorbance at 280 nm, black line). Molar masses (red dots) and hydrodynamic radii (blue dots) are superimposed to the SEC chromatogram and calculated online by multi-angle and dynamic light scattering, respectively.

To gain more insight into M- β Gal thermal robustness, we assayed the enzyme long-term stability at three different temperatures (5, 37, and 50 °C),

monitoring the decrease of both activity and structure over time (Fig. 2). Inactivation was determined on the base of residual enzyme activity and denaturation was

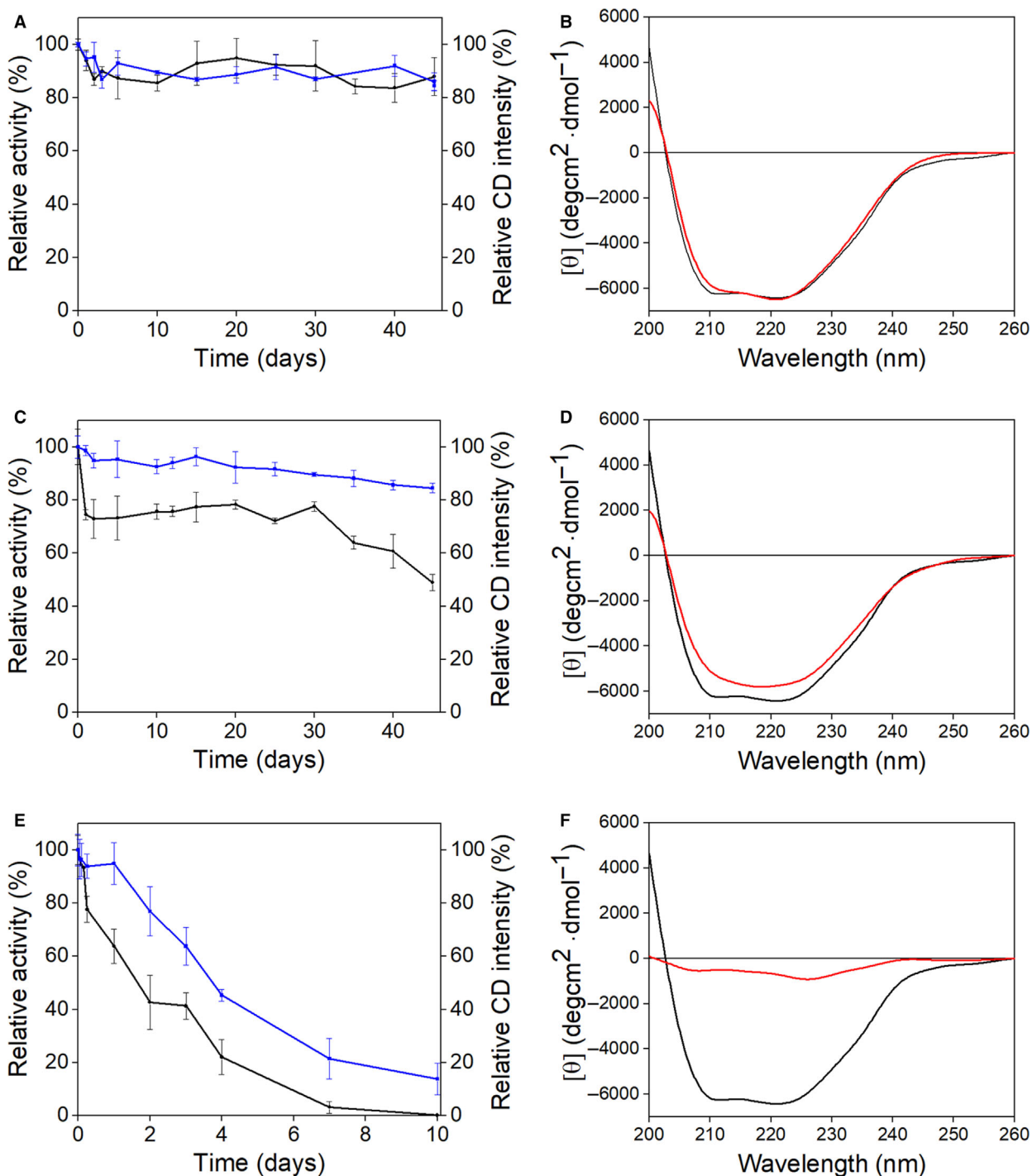


Fig. 2. Thermal stability of M- β Gal. Thermal stability over time was measured at 4 °C (A, B), 37 °C (C, D), and 50 °C (E, F) as residual activity and secondary structure. On the left, the residual activity (black line) and the relative CD intensity measured at 222 nm (blue line) were monitored day by day after incubating the enzyme at different temperatures. Error bars indicate standard deviations on three independent experiments. On the right, the Far UV CD spectra collected at the beginning (black line) and at the end (red line) of each kinetics.

measured as relative CD intensity at 222 nm. At day 0, the specific activity is 500 U·mg⁻¹ and the CD spectrum displays two minima peaks at 222 nm and at 210 nm which are characteristic of α/β structures (black line in Fig. 2B,D,F). At 5 °C the protein is stable, since the loss of activity during 45 days of experiment keeps within 10% of the initial value (Fig. 2A) and the CD spectrum recorded after 45 days is comparable with that of day 0 (Fig. 2A,B). At 37 °C, the activity decreases by 20% and 50% after 5 and 45 days, respectively (Fig. 2C). Noteworthy, over the whole incubation time the intensity of the CD signal lowers by 15% only (dashed line in Fig. 2C), while the overall CD spectrum undergoes slight modifications (Fig. 2D). At 50 °C, the activity drops to 25% in 4 days and is completely lost within a week. At this temperature, after the sharp decrease recorded in the first day, the loss of activity clearly parallels the loss of secondary structure (Fig. 2E), probably accompanied by protein aggregation, as witnessed by the appearance of visible protein precipitates in the tubes and by the flattening of the CD signal (Fig. 2F). Overall, when compared to other homologous psychrophilic enzymes, M-βGal reveals higher thermal robustness [7,8,43].

Temperature is an allosteric effector of M-βGal

Steady-state kinetics were carried out at four different temperatures (5, 25, 40, and 55 °C) using ONPG as the substrate, and data analyzed by fitting with the Hill equation (Table 1). At 5 °C, the kinetics is well fitted by a typical Michaelis–Menten curve, as indicated by the value of the Hill coefficient (*n*) close to 1 (Fig. 1E, Table 1). In the range 5–40 °C, the Hill coefficient slightly increases with temperature (Table 1). Interestingly, at 55 °C, we observed a sigmoid curve (Hill coefficient: 3.1, Fig. 1E) consistent with increased cooperativity. At this temperature, *V*_{max} (60.1 ± 1.7 mm·s⁻¹) undergoes a 12-fold increase compared with that observed at 5 °C (5.7 ± 1.0 mm·s⁻¹). By contrast, the affinity for the substrate (*K*_{0.5}) is higher at 5 °C (*K*_{0.5}: 6.9 ± 0.4 mM) than at 55 °C (*K*_{0.5}: 14.3 ± 0.7 mM).

Table 1. Kinetics parameters and Stern–Volmer quenching constant of M-βGal.

	<i>K</i> _{0.5} (mM)	<i>V</i> _{max} (mm·s ⁻¹)	<i>n</i>	<i>K</i> _{SV} (mm ⁻¹)
5 °C	6.9 ± 0.4	5.7 ± 1.0	1.3	2.4
25 °C	7.1 ± 0.5	26.9 ± 0.6	1.4	3.8
40 °C	9.7 ± 0.5	30.7 ± 0.9	1.5	4.7
55 °C	14.3 ± 0.7	60.1 ± 1.7	3.1	6.2

M-βGal is phylogenetically related to thermophilic GH42s

To frame the peculiar combination of properties shown by M-βGal (cold activity, high *T*_{opt}, high thermal stability) in the context of the evolutionary history of this enzyme, we performed a phylogenetic analysis and estimated divergence times among GH42 protein sequences in a data set composed by all characterized GH42s present in the CAZy database, and other GH42s from psychrophilic organisms (see **Materials and methods** section). The resulting rooted tree (Fig. 3) shows that sequences from psychrophilic (blue dots in figure), mesophilic (empty dots), and thermophilic (red dots) organisms do not cluster according to the organism natural environment. Instead, sequences from different types of extremophiles are closely related. Considering their occurrence in several different clades, it is reasonable to hypothesize that cold-adapted enzymes may have evolved independently from different ancestors at different times. M-βGal is closely related to the GH42 protein from the Antarctic *Marinomonas polaris* (uncharacterized) and both of them share with the cold-active *Marinomonas* sp. BSi20414 (MaBGA, [14]) an older ancestor dated around 40 millions years ago (mya). These three enzymes are nested in a group with eight sequences from thermophilic organisms, sharing a common ancestor of ≈ 200 mya with the sequences from *Thermus* species (Fig. 3). On this ground, we can hypothesize that the unusual biochemical properties of M-βGal and MaBGA may reflect both the adaptation to cold environments and common traits of thermophilic homologues of the same protein family.

The M-βGal structure is organized in three domains

The 3D structure of M-βGal was characterized by means of X-ray crystallography. M-βGal crystals belong to the hexagonal space group *P*6₃22, with one protein molecule in the asymmetric unit and a very high solvent content (~ 69.6%), corresponding to an unusual Matthews coefficient of ~ 4.0 Å³·Da⁻¹. The structure, refined to 1.9 Å resolution to final *R*-factor and *R*-free of 16.3% and 18.9% respectively, consists of the full-protein construct (residues 1–656), 2 Cl⁻ ions, 10 glycerol molecules originating from the cryoprotectant solution used prior X-ray data collection, and 978 water molecules (Table 2).

M-βGal has a three-domain (A, B, and C) tertiary structure (Fig. 4A), and a quaternary hexameric structure formed by a dimer of trimers (Fig. 5). In each

Table 2. Data collection and refinement statistics of M-βGal 3D structure.

Data collection	
X-Ray source	ESRF ID14-4
Space group	<i>P</i> 6 ₃ 22
Unit cell dimensions	
a, b, c (Å)	133.3, 133.3, 232.6
α, β, γ (°)	90, 90, 120
Wavelength (Å)	0.87313
Resolution (Å)	116.3–1.90 (2.00–1.90) ^a
Observations	2 830 997 (114 245)
Unique reflections	96 335 (3405)
<i>R</i> -merge (%) ^b	25.5 (67.4)
//σ(<i>I</i>)	10.9 (4.1)
Completeness (%)	100 (100)
Multiplicity	29.4 (25.6)
Refinement	
<i>R</i> -factor/ <i>R</i> -free (%) ^c	16.3/18.9
Protein residues in the a.u.	656
Cl [−] ions	2
Glycerols	10
Waters	978
Model quality	
<i>B</i> -factor (Å ²)	
Protein residues	19.1
Cl [−] ions	17.4
Glycerols	56.6
Waters	34.6
RMSD from ideal values	
Bond lengths (Å)	0.008
Bond angles (°)	1.25
Ramachandran plot	
Favored regions (%)	98
Allowed regions (%)	2
Outliers (%)	0

^aValues in parentheses refer to the last resolution shell.

^b $R\text{-merge} = \frac{\sum |I - \langle I \rangle|}{\sum I} \times 100$, where *I* is the intensity of a reflection and $\langle I \rangle$ is the average intensity.; ^c $R\text{-factor} = \frac{\sum |F_o - F_c|}{\sum F_o} \times 100$. *R*-free is calculated from 5% randomly selected data for cross-validation.

Geobacillus stearothermophilus (Gs-βGal) [30], and *Bifidobacterium animalis* subsp. *lactis* BL-04 (Bl-βGal) [29] (Table 3). The conservation is particularly marked when considering the isolated domains (root mean-square deviation—RMSD—ranges of 0.7–1.6 Å for the A domain, 1.1–1.8 Å for the B domain, and 1.2–1.6 Å for the C domain). A slightly different domain orientation does not affect the trimeric assembly (see next paragraph), which is conserved in GH42s and is relevant for the organization of the active site.

Most significant differences are found in the length and conformation of some local loops (Figs. 4B and 6) and in the coordination of Cl[−] ions. The first Cl[−] ion is coordinated by the side chains of Trp12 and Arg353 and the main chain NH of Tyr8 and Arg355 (Fig. 7A), and it superimposes to that found in A4-βGal. The second ion, not conserved in other GH42s, is bound to the NH of Val149 and Leu150 and to water molecules in a solvent-exposed surface cleft between the core of the α/β barrel and the H subdomain. Of note is that the coordination of the Cl[−] ion by M-βGal Arg353 and A4-βGal Arg351 forces the preceding Phe residue in a *cis* configuration in both proteins.

M-βGal shares the trimeric assembly of GH42 β-galactosidases

The first level of M-βGal protomer organization gives rise to a trimer, which is the lone quaternary structure of other GH42 β-galactosidases of known 3D structure [20,27–30]. The overall shape of the M-βGal trimer resembles an inverted ‘flowerpot’, with each protomer interacting with the other two (Fig. 5). The maximum internal diameter of this flowerpot is about 20–25 Å, the narrowest point (about 8 Å) being located at the

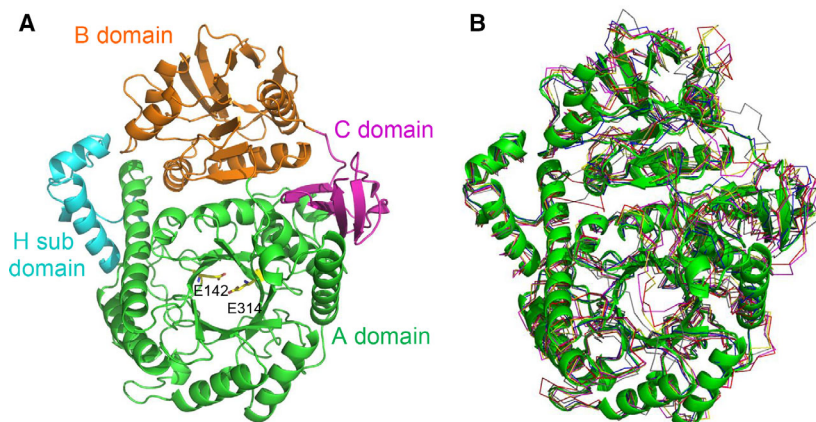


Fig. 4. Tertiary structure of M-βGal. (A) Ribbon diagram showing the 3D structure of M-βGal. The A, B, and C domains are shown in green (with the H subdomain in cyan), orange, and magenta. Catalytic residues Glu142 and Glu314 are labeled and shown as yellow sticks. (B) Structural superposition of M-βGal (green) with A4-βGal (PDB code 1KWG, blue [20]), Bc-βGal (PDB code 3TTS, red, [27]), R-βGal (PDB code 5E9A, yellow [28]), Gs-βGal (PDB code 4OIF, magenta, [30]), Bl-βGal (PDB code 4UNI, gray, [29]). The figure was prepared using Pymol software [87].

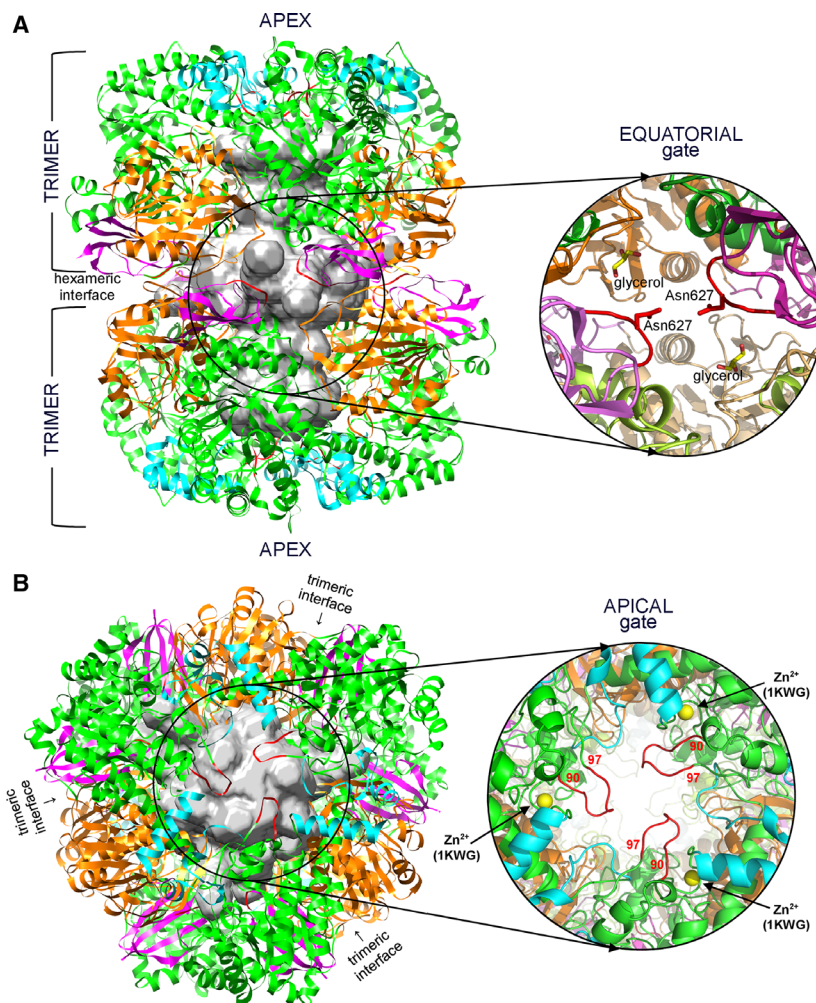


Fig. 5. Quaternary structure of M- β Gal. Ribbon representation of the hexameric structure (dimer of trimers) of M- β Gal. Domain coloring as in Fig. 4. (A) Side view, centered on one of the three symmetrical equatorial gates, and its close-up view with Asn627 residues and glycerol molecules highlighted (right panel). The internal cavity, inside the hexameric structure, is shown as water-excluded surface (gray, left panels in A and B). (B) Top view, centered on one of the two apical gates, and its close-up view (right panel). Flexible loops are shown in red, glycerol molecules as sticks. Zn^{2+} ions present in A4- β Gal (PDB code 1KWG) are shown as yellow balls, after structural superimposition of A4- β Gal with M- β Gal. The figure was prepared using CHIMERA software [88].

facing of the 90–97 loops from the three subunits, at the apex of the trimer (Fig. 5A). Within the trimer, each protomer interacts with two adjacent protomers via two main contact regions involving the domain A (C-terminal loops of the α/β barrel), and a second contact area formed by parts of domain B and subdomain H (Fig. 4B). Many residues participate in these intermolecular contacts, providing van der Waals interactions (93 residues), hydrogen bonds (49 residues), and salt bridges (5 residues). The trimeric assembly of M- β Gal buries about 4470 \AA^2 solvent-accessible surface area (SASA) in each protomer and it is predicted to be stable by the program PISA [46]. Unless some sequence variations at the inter-subunit regions, the overall trimerization interface is well conserved in various GH42 structures, despite their different crystallographic environments [20,27–30].

The three active sites of M- β Gal (Figs 4 and 5) are all facing the internal core of the trimer at the

interfaces between protomers. Electrostatic surface calculations reveal that the active site entrances are negatively charged (Fig. 7C), as frequently observed in other GHs [30]. More in detail, the catalytic residues of M- β Gal, Glu142 (acid/base) and Glu314 (nucleophile) are conserved and positioned at the C-terminal ends of the β_4 and β_7 strands, at $\sim 4\text{--}5 \text{ \AA}$ distance from each other (Figs 4A and 6). The active site of M- β Gal contains a glycerol and two water molecules and perfectly overlaps the active sites of three GH42 enzymes of known structure, in complex with a galactose molecule [20,27,29] (Fig. 7B). The three hydroxyl groups of the bound glycerol are at H-bond distance to the side chains of Glu362, Arg103, Trp322, Tyr267, His365, and the catalytic residue Glu142, while Arg33, Phe37, Asn141, Glu314, and Phe352 complete the lining of the active site cavity. The M- β Gal active site is completed by the presence of a highly conserved Trp183, which is involved in substrate binding and/or

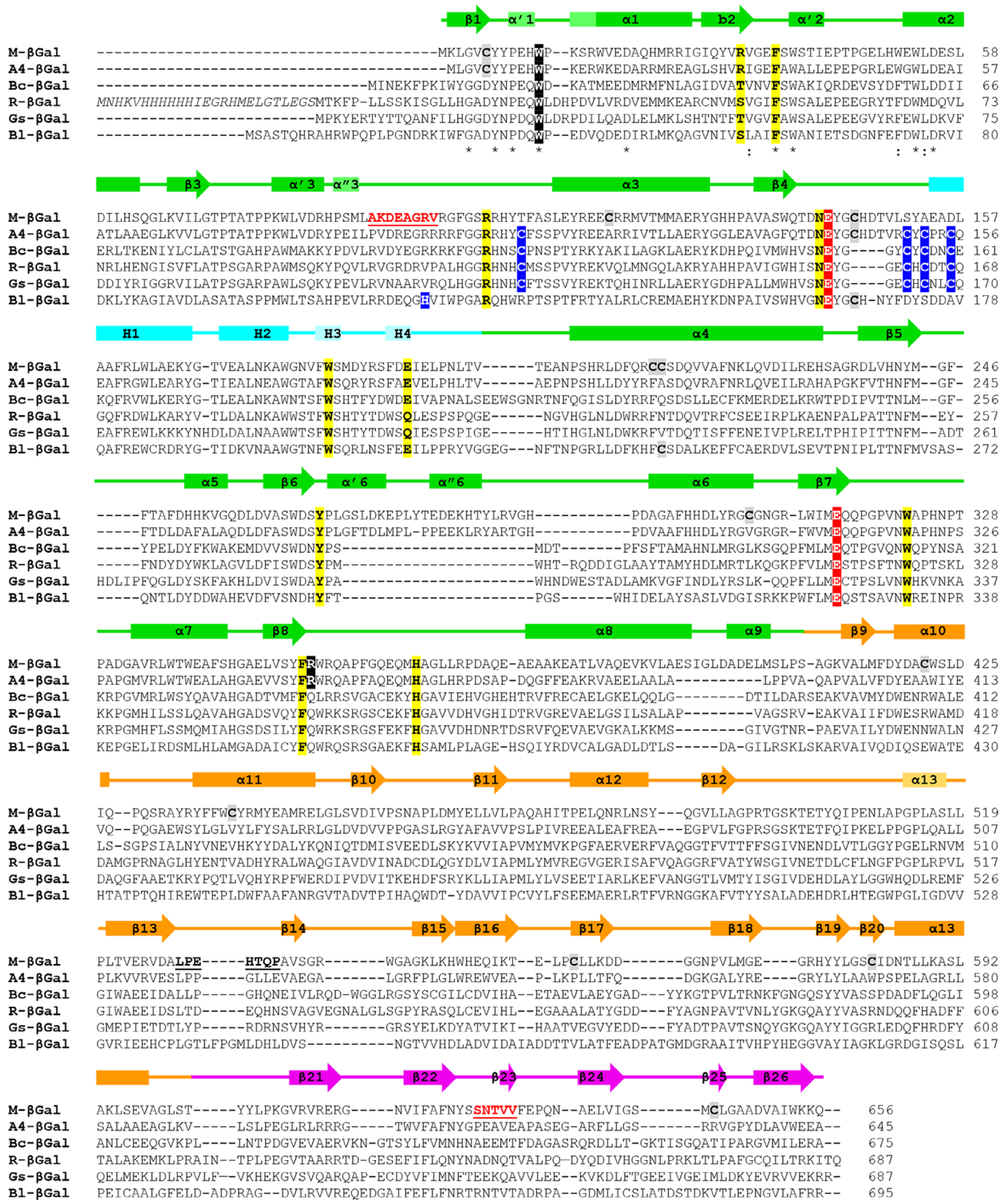


Fig. 6. Structure-based sequence alignment of M- β Gal with related GH42s. M- β Gal is aligned with β -Gal from *Thermus thermophilus* A4 (A4- β Gal, PDB code 1KWG, [20]), *Bacillus circulans* sp. (Bc- β Gal, PDB code 3TTS, [27]), *Rahnella* sp. R3 (R- β Gal, PDB code 5E9A, [28]), *Geobacillus stearothermophilus* (Gs- β Gal, PDB code 4OIF, [30]), and *Bifidobacterium animalis* subsp. *lactis* BL-04 (Bl- β Gal, PDB code 4UNI, [29]). Protein sequences were retrieved from the Protein Data Bank. Sequence alignment was performed using the CLUSTALW program (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and manually corrected based on 3D structure comparison. M- β Gal secondary structure elements are shown at the top of the alignment (color code as in Fig. 4, with 3_{10} helices in a lighter color). Catalytic residues are shaded in red and residues lining the active site in yellow. M- β Gal Cys residues and the corresponding Cys residues in GH42 aligned enzymes are shaded in gray. Cys residues forming the Zn²⁺ binding site in GH42 proteins and the corresponding His residue in Bl- β Gal are shaded in blue. M- β Gal Trp12 and Arg353 involved in Cl⁻ coordination are shaded in black. Residues delimiting the access gates to the M- β Gal hexameric cavity, at the apex and in the equatorial regions of the hexamer, are in bold character and underlined (in red the flexible loops). When present, residues belonging to the purification tag are in italic.

Table 3. Structural comparison of M- β Gal with homologues GH42 enzymes.

	M- β Gal (6Y2K)	R- β Gal (5E9A)	Bc- β Gal (3TTS)	Bl- β Gal (4UNI)	A4- β Gal (1KWG)	Gs- β Gal (4OIF)
N ^o of residues	656 (1–656)	684 (2–685)	675 (1–675)	689 (7–695)	644 (1–644)	685 (3–687)
Molecular mass ^a (Da)	74 332.16	77 245.19	77 327.50	78 241.45	72 824.08	79 992.42
Secondary structure ^b						
Helix	35% (234) ^c	30% (208)	31% (212)	30% (213)	34% (222)	32% (223)
Beta	27% (183)	32% (218)	31% (212)	30% (212)	28% (183)	32% (221)
Turn	25% (164)	20% (136)	21% (148)	24% (168)	24% (160)	21% (148)
Coil	36% (239)	37% (253)	37% (251)	38% (264)	37% (239)	35% (241)
Fraction of Gly ^a	7.2%	8.0%	7.3%	6.5%	8.2%	5.7%
Fraction of Pro ^a	5.8%	4.8%	4.0%	5.0%	8.5%	4.1%
Fraction of Arg ^a	5.8%	5.5%	5.3%	6.5%	8.8%	5.8%
Total volume ^b (Å ³)	93 664.7	97 316.1	100 477.4	95 472.4	95 230.4	99 117.0
MeanRes volume ^b (Å ³)	142.8	143.3	148.9	138.6	147.9	144.7
Packing defects ^b	141	160	211	90	219	80
Res 95% buried ^b	254	274	265	263	257	270
Buried charges ^b	20	25	19	24	22	24
Total ASA ^b (Å ²)	25 228.3	25 651.0	25 671.0	26 073.1	24 415.7	26 353.7
ASA backbone ^b (Å ²)	2758.5	3113.1	2953.5	3150.7	2913.9	2896.7
ASA sidechain ^b (Å ²)	22 469.8	22 537.9	22 717.6	22 922.4	21 501.8	23 457.0
Fraction nonpolar ASA ^b	0.58	0.59	0.58	0.57	0.60	0.57
Fraction polar ASA ^b	0.19	0.21	0.20	0.19	0.16	0.18
Fraction charged ASA ^b	0.23	0.20	0.22	0.25	0.24	0.24
Intra-protein interactions ^d						
Van der Waals						
All	696	743	772	754	703	807
MC-MC	2	1	1	1	1	1
MC-SC	127	176	165	152	120	164
SC-SC	567	566	606	601	582	642
Hydrogen-bonds						
MC-MC	494	466	486	475	471	491
MC-SC	87	95	98	111	63	102
SC-SC	27	35	34	47	33	40
Ionic	36	24	28	33	30	35
π - π stacking	59	46	39	39	59	58
DALI analysis						
DALI Z-score	–	40.1	40.5	38.4	52.8	38.4
Residue identity (%)	–	28	29	26	55	26
RMSD (Å)	–	2.1	2.0	2.5	1.0	2.2
Reference	This study	[28]	[27]	[29]	[20]	[30]

^aCalculated with ProtParam [84].; ^bCalculated with Vadar [85].; ^cThe N^o of residues is in parentheses.; ^dCalculated with RING [86].

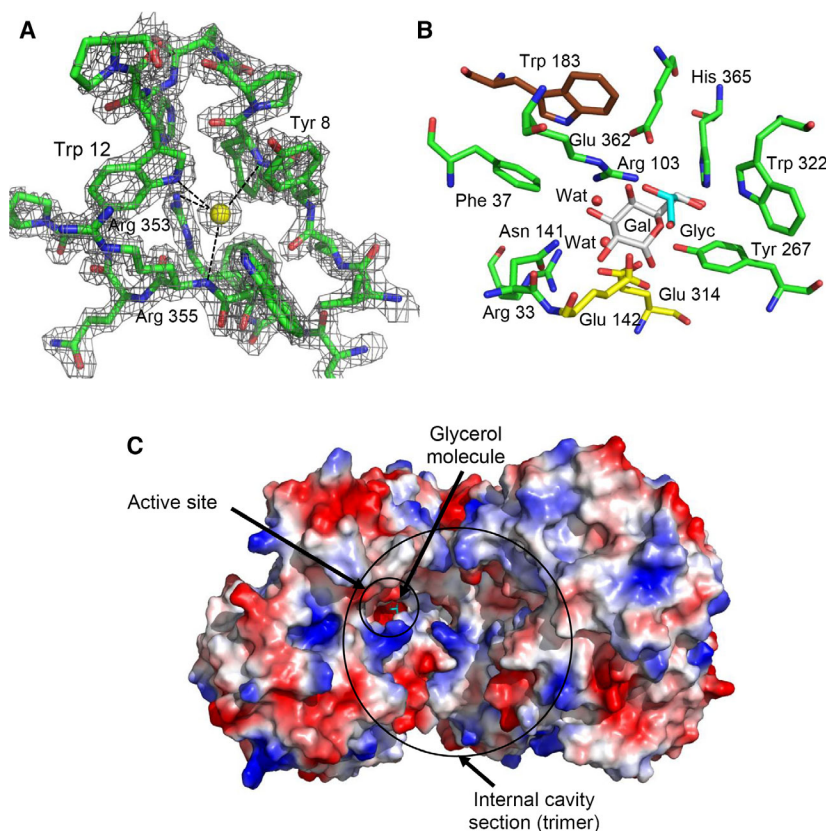


Fig. 7. M- β Gal binding sites. (A) Structure of the protein around the Cl^- binding site. The 2Ifol-IFcII electron density around the ion-binding site is shown as a gray mesh. Residues involved in ion coordination are labelled. (B) Active site of M- β Gal in complex with glycerol (cyan). Catalytic residues Glu142 and Glu314 are shown as yellow sticks, the other residues lining the active site as green sticks, and water molecules as red spheres. Trp183, which belongs to an adjacent subunit, is shown in brown. A galactose molecule (gray sticks), derived by the complex structure with Bc- β Gal (PDB code 3TTY, [27]), is shown in the M- β Gal active site, after superimposing M- β Gal with Bc- β Gal. (C) Electrostatic surface representation of the active site of one M- β Gal subunit of the hexamer (blue positive, red negative charges). For clarity, only the protein surface of two adjacent protomers of one trimer is shown. The absence of the front subunit of the trimer allows viewing the internal cavity. Overall, six identical active sites face the internal hexameric cavity. The figure was prepared using PYMOL software [87].

recognition [30]. This residue belongs to the adjacent symmetry-related protomer, which is part of the trimeric quaternary structure.

The quaternary hexameric structure and the absence of Zn^{2+} binding sites are peculiarities of M- β Gal

The hexameric arrangement of M- β Gal is unprecedented in the GH42 family. Size-exclusion chromatography combined with light scattering detection (SEC-LS) analysis confirms crystallographic data, indicating that the dimension of the native enzyme in solution (molar mass 396 kDa, R_h 64 Å) (Fig. 1F) is compatible with a hexameric assembly.

In M- β Gal, the hexameric arrangement is the result of the dimerization of two trimeric flowerpots. Dimeric assembly of trimers buries about 930 Å² in each protomer and involves residues located only in the B and C domains (Fig. 5A), which form the ‘base’ of the flowerpot. In the hexameric structure, six active sites face a very large internal cavity of about 54000 Å³, having a threefold symmetry along the vertical axis of the hexamer (Fig. 5). Some trimer–trimer association (albeit different from M- β Gal) was also reported for

A4- β Gal, R- β Gal, and BI- β Gal, but in all these cases, the small area of the buried interface and SEC data clearly indicate that this association is the result of crystal packing [27–29]. The access/exit to the internal cavity could occur through two different types of gates, namely ‘equatorial’ and ‘apical’ gates. The three equatorial gates derive from packing defects at the trimer–trimer interface, close to the β 13- β 14 loop (529–535 residues) and to the β 22- β 23 loop (626–630 residues), and were localized by the presence, inside the cavity, of two glycerol molecules from the cryoprotectant solution (Fig. 5A, right panel). Equatorial gates are located at the end of a deep surface cleft at the interface of protomers and lined by mostly polar/charged residues, such as Asp331, Ser626, Asn627, T628, Ser642, and Asp659. The cleft has a narrowing in correspondence of Asn627, which splits the gate in two almost circular parts of about 10 Å of diameter each (Fig. 5A). Besides equatorial gates, the top of the two flowerpots (loop 89–100) might also offer another access to the internal cavity (Fig. 5B, right panel). These two apical gates, of about 8 Å in diameter, are located at the end of a surface funnel, generated by three interfacing protomers, and lined mostly by polar/charged residues, such as Ser87, Lys91, Arg96, Asp92, Glu154, and Glu194.

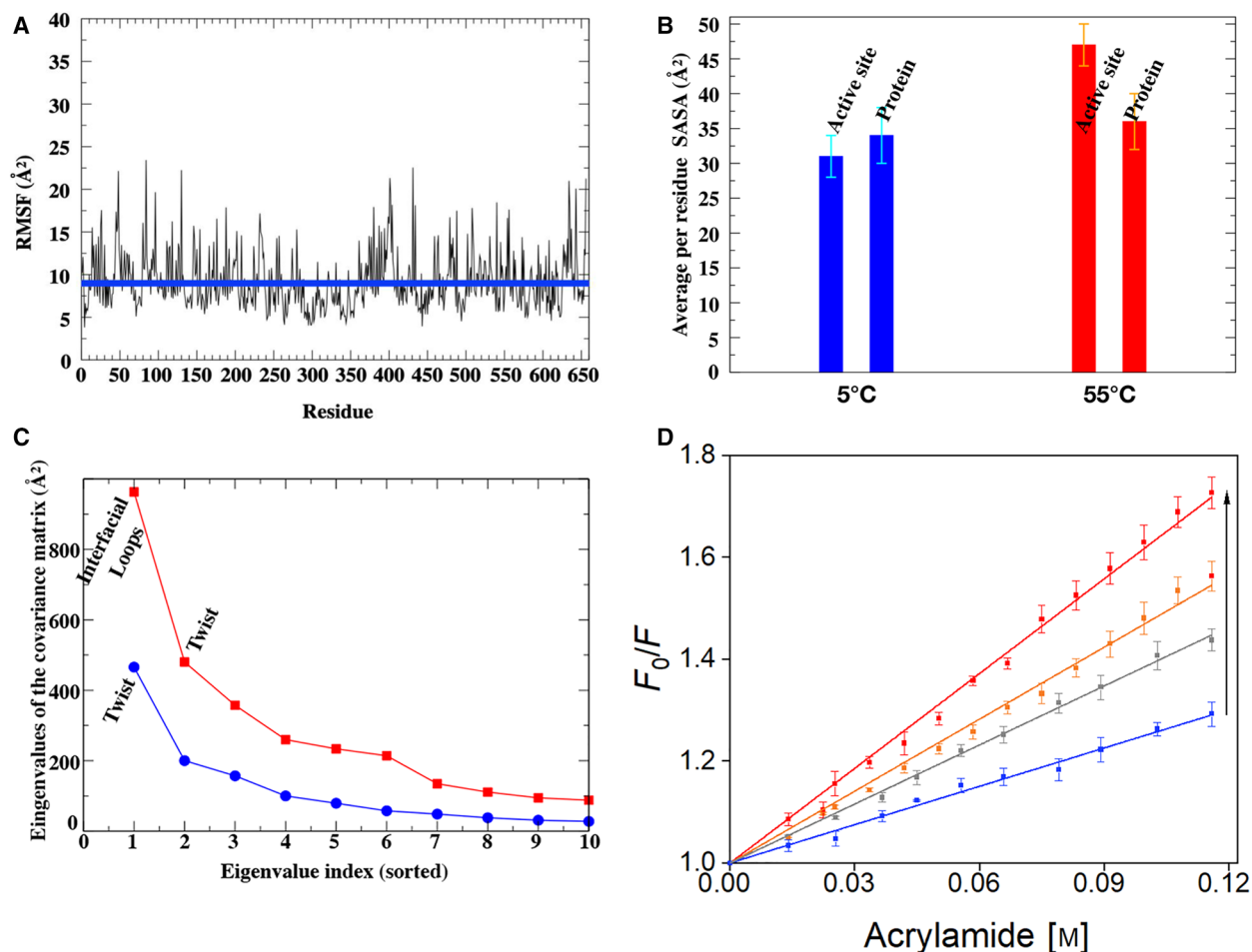


Fig. 8. Dynamics properties of M- β Gal. (A) Per-residue root-mean-square fluctuations (RMSF) (\AA^2) averaged over the hexamer for the MD simulation at 5 $^{\circ}\text{C}$. The blue line at 9.0 \AA^2 is the average value. (B) Per-residue solvent-accessible surface area (SASA) (\AA^2) averaged over the hexamer. Blue bars refer to the MD simulation at 5 $^{\circ}\text{C}$, while red bars to the MD simulation at 55 $^{\circ}\text{C}$. Error bars report the standard deviation calculated over all the hexamer residues. (C) Sorted eigenvalues (first ten) resulting from the diagonalization of the covariance matrix for the C_{α} carbons of the hexamer. Blue circles refer to the MD simulation at 5 $^{\circ}\text{C}$, while red squares are for the MD simulation at 55 $^{\circ}\text{C}$. The motion associated with the first eigenvalue is annotated in the plot. (D) Stern–Volmer plots of fluorescence quenching by acrylamide monitored at 5, 25, 40, and 55 $^{\circ}\text{C}$. Color code is the same as in (1E). The arrow on the right indicates increasing temperature. Error bars indicate standard deviations on three independent experiments.

A further distinctive feature of M- β Gal is the lack of the Zn^{2+} binding site conserved in all other GH42s with known 3D structure, where Zn^{2+} is coordinated at the top of the trimer either by four Cys [20,27,28,30], or by a water molecule and three His residues, one from each monomer [29]. The absence of the Zn^{2+} binding site might increase the flexibility of the apical loops of M- β Gal, favoring the entrance of substrate molecules through the apical gates (Fig. 5B). This hypothesis was examined by molecular dynamics (MD) simulations carried out on the M- β Gal hexamer at 5 $^{\circ}\text{C}$. At this temperature, all-atom RMSD of each

protomer is within 2 \AA , and the average radius of gyration (R_g) of the hexamer is 48.6 \AA , R_h of 61.0 \AA [47], in agreement with the R_g and R_h obtained by X-ray structure and SEC-LS (47.4 and 64.0 \AA , respectively). The analysis of fluctuations of the atomic positions suggests that the apical loops at the entrance of the large cavity are significantly more flexible (root-mean-square fluctuation, RMSF 12 \AA^2) than the average of the protein residues (RMSF 12 \AA^2). Furthermore, also the β 22- β 23 loop at the equatorial gates seems to be more flexible than the other loops located at the subunit–subunit interfaces (RMSF 9.5 \AA^2) (Fig. 8A).

Heating increases the accessibility of the active sites

To explore more in depth how temperature affects the accessibility of the active sites, an additional simulation was performed at 55 °C. The solvent-accessible surface area (SASA) was calculated at 5 and 55 °C in the range from 200 to 350 ns, when a relative stabilization of M- β Gal was reached. As expected, the overall SASA increases more at higher temperature (15% vs 20% at 5 and 55 °C, respectively). To note, at 5 °C the average SASA per active site residue (31 Å²) is very similar to that observed for all other residues (34 Å²). Upon heating at 55 °C, active site SASA increases significantly (47 Å² vs 36 Å² observed for the average of all other residues), suggesting that specific rearrangements affect mainly the active sites. To gain more detail about protein motions, we performed a covariance analysis on the C α carbons of the hexamer. The trend of the sorted eigenvalues is comparable at both temperatures, with the difference in absolute value simply associated with the larger fluctuations at higher temperature. At 5 °C, the first eigenvector is associated with a twisting motion of the two trimers within the hexamer, which slightly rotate one with respect to the other without affecting the accessibility of the active sites that move rigidly within each trimer (Fig. 8B). On the contrary, at 55 °C the first eigenvector is associated with motions of loops at the interface between protomers in each trimer, and the loop including the Trp183 that is in close contact with the active site of the symmetric protomer (Fig. 7B). Such motions make the active site of each hexamer protomer more accessible to the solvent and hence to the substrate. Of note, the second eigenvector at 55 °C corresponds to the first at 5 °C (Fig. 8B). Altogether, our MD simulations suggest that, upon heating, active sites become more accessible *via* hexamer correlated motions.

The overall structural compactness of M- β Gal was further explored by dynamic fluorescence quenching by acrylamide of tryptophan residues (21 in M- β Gal), at 5, 25, 40, and 55 °C. In this kind of experiment, the diffusion of acrylamide is limited by the protein structural compactness, and the quenching extent varies with acrylamide concentration at a rate described by the Stern–Vollmer constant (K_{SV}), which is also temperature-dependent. K_{SV} values were calculated through the linear fitting of the curves in Fig. 8D, according to Equation 2 (see [Materials and methods](#) section). Overall, data obtained for M- β Gal (Table 1) indicate that heating induces greater accessibility to acrylamide of Trp residues buried in the M- β Gal structure.

Discussion

M- β Gal is an atypical cold-active enzyme

M- β Gal conforms to the current definition that classifies as cold-active those enzymes able to maintain activity at low temperature, typically 10–20% of their highest activity at ≤ 10 °C [7,10,11]. The T_{opt} at 55 °C, although higher than that of most cold-active enzymes, is not a rarity among psychrophilic GH42s [14,32,33,36]. More astonishing is the heat resistance of M- β Gal, which has T_m of 68.7 °C and remains active after several days at 37 °C and a few days at 50 °C. This observation is very remarkable as at the latter temperature most psychrophilic enzymes have half-lives of a few minutes only [10,48,49]. Thus, M- β Gal couples apparently mutually exclusive properties such as cold activity and high thermostability. In seminal studies, it has been shown that contrasting properties can be combined through directed-evolution experiments when appropriate multiple selective pressure is applied [50]. Thus, for example, the use of directed-evolution approaches and multiple screening allowed developing cold activity in mesophilic enzymes and *vice versa*, without losing the original temperature dependence [51]. Our phylogenetic study of GH42 sequences suggests that M- β Gal may have arisen as a rather thermostable protein scaffold on which cold activity was implanted as a consequence of changed environmental conditions experienced by the producing bacteria.

Cold adaptation of M- β Gal is based on unexpected structural solutions

Cold-active enzymes developed different structural changes to increase the flexibility of the protein surface or active site (*e.g.*, peculiar amino acids composition, lower number of weak interactions). The comparison of M- β Gal with other GH42s indicates that these enzymes are similar in sequence and structural features (Table 3). Although some differences in the length and hydrophobicity of connecting loops were identified (Fig. 6), they are distributed in the GH42 family without any specific rationale related to temperature adaptation. Whenever adaptation strategies involve the protein topology or oligomeric state [52–54], the quaternary structure of psychrophilic proteins often displays more flexible oligomerization interfaces relative to their mesophilic and thermophilic homologues [53,55]. Therefore, we hypothesize that cold adaptation and the ‘hybrid’ functional behavior of M- β Gal are related to its unique hexameric arrangement. We

found that the quaternary assembly of M- β Gal buries ($\sim 54\,000\ \text{\AA}^3$), hosting 1460 ordered water molecules, 6 Cl^- ions, and 12 glycerol molecules (in the first hydration shell of the protein protomers). The volume of the internal cavity is ~ 1.7 -fold larger than the reaction chamber of propionyl-CoA synthase from *Erythrobacter* sp. NAP1 ($\sim 33\,000\ \text{\AA}^3$) [56] and is similar to that of the proteasome antechamber ($\sim 59\,000\ \text{\AA}^3$) [57]. The presence of such a large internal cavity containing solvent molecules is considered as a mean to increase the structural flexibility of psychrophilic proteins [58]. We hypothesize that such a wide substrate-binding chamber could favor the binding events at cold temperatures and counteract the low substrate affinity typical of cold-active enzymes.

Five 'gates' (two at the apices and three at the equatorial region of the hexamer) connect the internal substrate chamber with the external bulk solution. Overall, the movements of the loops belonging to all five gates contribute the most to the dynamic of the protein surface, as witnessed by MD simulations at $5\ ^\circ\text{C}$. In particular, the absence of a Zn^{2+} binding site, otherwise conserved in GH42s, would confer flexibility to the apical gates (Figs 6 and 8A). We propose that such a combination of flexible elements coupled with the hexameric arrangement found in M- β Gal can represent a novel cold adaptation mechanism, unshared by trimeric psychrophilic GH42s, such as R- β Gal, in which active sites are fully exposed to the solvent, and therefore may conserve the Zn^{2+} binding site at the top of each trimer [28].

Heating modulates substrate affinity and triggers binding cooperativity

Substrate-binding affinity is reported as a key factor in temperature adaptation [59,60]. Usually, substrate-binding sites of cold-active enzymes are larger than those of their mesophilic and thermophilic homologues, resulting in weaker substrate-enzyme interactions and lower binding affinity. Overall, the enlargement of substrate-binding sites lowers the free energy barrier of the transition state, which ends up fostering catalysis [60]. As a consequence, the K_M of several psychrophilic enzymes is higher compared to those of mesophilic counterparts [8,43,59–61]. However, the $K_{0.5}$ of M- β Gal, determined at $5\ ^\circ\text{C}$, is similar to those of mesophilic and thermophilic GH42s [29–31,35–37,62–64], pinpointing the hexameric structure, with its large internal cavity, as an elegant and unusual strategy to modulate substrate accessibility and affinity. Heating increases $K_{0.5}$, V_{max} , and cooper-

ativity, indicating that temperature could act as an allosteric effector for M- β Gal. Such an unexpected kinetic behavior can be accounted for by the greater flexibility of chamber gates, the faster substrate diffusion, and the further enlargement of the chamber volume triggered by heating, as suggested by MD simulations.

In conclusion, our study contributes to sample the diversity of cold adaptation strategies within the members of a protein family exposed to similar selective pressure. This heterogeneity, probably due to the phylogenetic distance of the GH42 ancestors, leads us to consider the search for a common 'structural signature' of cold adaptation as utopian. Rather, cold adaptation appears to pursue the improvement of kinetics efficiency at low temperature through very diverse and, in some cases, unexpected structural solutions.

Materials and methods

Identification of M- β Gal and cloning

The sequence encoding for a putative β -galactosidase was identified in the *Marinomonas* efl genome deposited in the GenBank under the accession number NHHTT01000000 by the pipeline available in Prokka 1.12 [65] and by a Standalone Blast operation using the β -galactosidase from *Marinomonas polaris* (acc. N. WP072839227) as the query. The identified *Marinomonas* efl β -galactosidase (M- β Gal) is available in the GenBank database under the accession number WP100635792.

The synthetic gene encoding for M- β Gal was sequence-optimized for expression in *Escherichia coli* (GenScript, Piscataway, NJ, USA), and cloned in frame with a C-terminal 6xHis-Tag into pET-21a vector (EMD, Millipore, Billerica, MA, USA) between *Nde*I and *Xho*I sites to obtain pET-21 [M- β Gal].

Recombinant protein production and purification

Escherichia coli strain DH5 α^{TM} (Invitrogen, Waltham, MA, USA) was the host for plasmid DNA amplification, while strain BL21 (DE3) (EMD Millipore) was used for recombinant protein expression. M- β Gal was intracellularly produced in Zym 5052 [66] added of $100\ \text{mg}\cdot\text{L}^{-1}$ of ampicillin (Merck, Darmstadt, Germany) and extracted as described in [52]. Proteins were purified on gravity flow chromatography column of nickel-nitrilotriacetic acid agarose resin (Jena Bioscience, Jena, Germany). Samples containing the highest amount of protein were pooled and buffer-exchanged twice by gel filtration on PD10 columns (GE healthcare, Little Chalfont, UK) against 10 mM sodium phosphate buffer (PB), pH 6.0.

Protein concentration was determined by the Bradford protein assay (Bio-Rad, Hercules, CA, USA), using bovine serum albumin as a standard.

Activity assay

Enzyme activity was measured by a discontinuous assay using ONPG (Merck) as a substrate. The assay was performed at 25 °C in PB, after 3 min the reaction was stopped by adding an equal volume of 1 M sodium carbonate buffer, pH 11.0, and the absorbance detected at 420 nm (molar extinction coefficient: 4.6 mm⁻¹·cm⁻¹) with a Jasco V-630 UV/VIS spectrophotometer (JASCO Europe, Lecco, Italy). One unit of enzyme activity was defined as the amount of the enzyme catalyzing the formation of 1 μ mol of ONPG per minute.

The pH optimum was explored by measuring the activity in PB, in the pH range 4.0–10.0. The catalytic temperature optimum was explored in PB, pH 6.0, in the temperature range 5–85 °C. Thermal stability was assessed by measuring the residual activity at 25 °C after incubating the enzyme (500 U) in PB, pH 6.0 at 5 °C for 45 days, 37 °C for 45 days and 50 °C for 10 days.

The kinetics parameters of M- β Gal were measured using different concentrations of ONPG (from 1 to 20 mM), at the optimum pH and at four different temperatures (5, 25, 40, and 55 °C). Kinetics parameters were calculated with the ORIGINLAB software (OriginLab Corporation, Northampton, MA, USA), using the nonlinear fitting of Hill equation 1

$$V = \frac{V_{\max} \cdot [S]^n}{(K_{0.5})^n + [S]^n} \quad (1)$$

where V is the reaction rate, V_{\max} is the maximum reaction velocity, $[S]$ is the substrate concentration, $K_{0.5}$ is the substrate concentration giving rise to a reaction velocity that is 50% of V_{\max} , and n is the Hill coefficient.

CD spectroscopy

CD spectra of M- β Gal (1 μ M in PB) were collected in a J-815 spectropolarimeter (JASCO Europe), using a 0.1-cm path length cuvette. Measurements were performed with 0.1-nm data pitch and 10-nm·min⁻¹ scanning speed. All spectra were corrected for buffer contribution, averaged from four independent acquisitions, and smoothed by using a third-order least-square polynomial fit.

Thermal denaturation ramps were obtained measuring the variation of CD signal at 205 nm when progressively heating the sample from 5 to 90 °C. Measurements were performed in triplicate with a data pitch of 2 °C and temperature slope of 0.5 °C·min⁻¹.

Structural changes occurring during the thermal stability assay were assessed by monitoring the CD signal at 222 nm

after incubating the enzyme in PB, pH 6.0 5 °C for 45 days, 37 °C for 45 days, and 50 °C for 10 days. CD spectra were recorded in the range 200–260 nm at the beginning and at the end of experiments.

Dynamic quenching of fluorescence

The acrylamide-dependent quenching of intrinsic protein fluorescence was monitored at four different temperatures (5, 25, 40, and 55 °C) in the presence of 5 μ M of M- β Gal in a 2-mL initial volume of PB, after consecutive additions of 2.5 μ L (at 40 and 55 °C) or 5 μ L (at 25 °C) of 5.6 M acrylamide (Bio-Rad).

Fluorescence emission spectra were collected after excitation at 280 nm with a Cary Eclipse Spectrofluorimeter (Varian Australia Pty Ltd, Mulgrave, Vic., Australia), using quartz cuvettes of 1-cm path length. Slit width was 5 nm, both for excitation and emission. Spectra were acquired at 1-nm intervals, at 600 nm·min⁻¹ rate and 0.1-s averaging time.

The Stern–Volmer quenching constant K_{SV} , which is a measure of the accessibility of the fluorophore to the quencher, was calculated at each temperature according to the relation [67] shown in equation 2,

$$\frac{F_0}{F} = 1 + K_{SV} \cdot [Q] \quad (2)$$

where F_0 and F are the fluorescence intensity measured at 335 nm in the absence and in the presence of molar concentration of the quencher Q .

Size-exclusion chromatography with multi-angle and dynamic light scattering (SEC-LS)

SEC-LS detection was carried out using a Superose 12 column (10/300 GL, GE Healthcare), fitted to a high-performance liquid chromatography (HPLC) system composed by a Waters 515 HPLC Pump, a Waters 2487 Dual λ Absorbance detector (Waters, Sesto San Giovanni, Italy), a Wyatt Dawn Heleos multi-angle light scattering equipped with a dynamic light scattering, and a Wyatt Optilab T-rEX differential refractive index detector (Wyatt Technology, Santa Barbara, CA, USA).

200 μ L of 1 mg·mL⁻¹ samples was run in 100 mM PB, pH 6.0, at flow rate of 0.5 mL·min⁻¹. Molar mass and R_h were calculated by means of the ASTRA software (Wyatt), using a dn/dc value of 0.185.

Crystallization, structure determination, and refinement

Crystallization of M- β Gal was achieved by using the hanging drop vapor diffusion method on 24-well Linbro™ plates. Stick-like crystals were obtained after 3 weeks at

21 °C, by equilibrating the protein solution (6 mg·mL⁻¹) against 0.5 M NaCl, 100 mM Na-citrate (pH 5.6), and 2% ethylene imine polymer. They diffracted up to 1.9 Å resolution, using synchrotron radiation at 100 K (ESRF, Grenoble, France). Crystals belong to the hexagonal *P*6₃22 space group with one protein molecule per asymmetric unit (estimated solvent content 69.6%) (Table 2). XDS [68], SCALA [69], and the CCP4 suite programs [69] were used for reducing, scaling, and analyzing all the collected data.

M- β Gal structure was solved by molecular replacement with Phaser [70], using the structure of *Thermus thermophilus* A4 β -galactosidase (PDB code 1KWG) as the search model. The M- β Gal molecular model was checked manually with COOT [71] and refined with REFMAC5 [72] to the maximum resolution (Table 2). The stereochemical quality was assessed with MolProbity [73], and the quaternary assemblies were identified with PISA [46]. Finally, the atomic coordinates and the structure factors have been deposited in the Protein Data Bank (www.rcsb.org) with entry code 6Y2K.

MD simulations

MD simulations were performed with GROMACS 2018 [74] using the Amber03 force-field with scaled protein–water interactions and the TIP4P/2005 water model (Amber03ws) [75,76]. Starting from the 3D structure of the M- β Gal hexamer, the protein was first energy-minimized in vacuum retaining two Cl⁻ atoms per protomer while removing the crystallographic water molecules. The protein was then solvated in a box of 2933 nm³ of volume with 78 704 water molecules and 132 Na⁺ counter ions to neutralize the overall charge and additional ions corresponding to 0.1 M NaCl. The global system consisting of 377 000 atoms was eventually further energy-minimized. The solvated protein was thermalized at either 278 or 328 K, first keeping the protein atoms fixed in their energy-minimized positions and successively letting all the atoms move and the box size adapting to a pressure of 1 bar. Conventional MD simulations, with hydrogen mass repartitioning that allows using a time step of 4 fs [77], were performed at the two temperatures, at constant volume, for 350 ns.

Phylogenetic and sequence analysis

The phylogenetic tree and the multi-sequence alignment were performed with Bali-Phy 3.2 [78]. Besides the sequence of M- β Gal, the data set included 62 sequences of characterized GH42s present in the CAZy database, 3 sequences of uncharacterized GH42s from psychrophilic organisms, a GH42 sequence from *Marinomonas parvialis* (the same used for the Blast search, see above), and one from the Arctic *Marinomonas* sp. BSi20414 (MaBGA) [14]. Six independent analyses were performed using the default parameters, with an empirical substitution rate matrix (LG model) and a

rs07 insertion/deletion model [78,79]. The convergence and mixing of sample parameters were checked according to the manual (http://www.bali-phy.org/README.html#mixing_and_convergence). In order to remove the background noise at the beginning of the run, the first 50% of steps were discarded. The position of the root was inferred with a parsimony-based approach, minimizing the costs of duplication, transfer, loss events under a defined species phylogeny. This analysis was performed with RANGER 2 [80], using the previously obtained unrooted phylogeny and a chronogram tree in which each branch represents the evolutionary time. This tree was generated through the ‘Time Tree of Life’ website (<http://www.timetree.org/>). Three different cost combinations for duplication (D), transfer (T), and loss (L) events were used, in order to select the optimal root position that minimizes the sum of costs for all the events in the tree. Each analysis was repeated 100 times for consistency. In order to estimate the posterior distribution of divergence times, the obtained rooted tree topology and the maximum *a posteriori* alignment were used as the input in two independent Bayesian runs of \approx 30 000 cycles in Phylobayes 4.1c [81]; the first 50% of cycles were discarded after verifying convergence and good mixing through *trace-comp* command. The CAT infinite mixture model [82], a log normal (Brownian) autocorrelated clock [83], and a Dirichlet prior on branching times (hyperprior parameter set to 0.5) were used along the same model specifications previously adopted. Four nodes connecting orthologous sequences were calibrated with hard but permissive symmetric boundaries around the divergence times for the corresponding species reported in the ‘Time Tree of Life’ website.

Acknowledgements

MM acknowledges support by a post doc research fellow (Assegno di Ricerca) of the University of Milano-Bicocca.

Conflict of interest

The authors declare no conflicts of interest.

Authors contribution

MM performed biochemical and biophysical characterization; MO analyzed sequence evolution; SM developed production and purification of the recombinant protein; AP crystallized the protein; ML and MN solved, refined, and analyzed the 3D structure; CC performed MD simulations; AB performed SEC-LS experiments; SB supervised the project and reviewed the manuscript; SP performed genome analysis and designed the synthetic gene; ML and MN conceived

the project and wrote the manuscript. All authors have read and approved the manuscript.

References

- Mangiagalli M, Brocca S, Orlando M & Lotti M (2020) The “cold revolution” Present and future applications of cold-active enzymes and ice-binding proteins. *New Biotechnol* **55**, 5–11.
- Sarmiento F, Peralta R & Blamey JM (2015) Cold and hot extremozymes: industrial relevance and current trends. *Front Bioeng Biotechnol* **3**, 1–15.
- Elleuche S, Schroeder C, Sahm K & Antranikian G (2014) Extremozymes—biocatalysts with unique properties from extremophilic microorganisms. *Curr Opin Biotechnol* **29**, 116–123.
- Collins T & Margesin R (2019) Psychrophilic lifestyles: mechanisms of adaptation and biotechnological tools. *Appl Microbiol Biotechnol* **103**, 2857–2871.
- Bruno S, Coppola D, di Prisco G, Giordano D & Verde C (2019) Enzymes from marine polar regions and their biotechnological applications. *Mar Drugs* **17**, 544.
- De Maayer P, Anderson D, Cary C & Cowan DA (2014) Some like it cold: understanding the survival strategies of psychrophiles. *EMBO Rep* **15**, 508–517.
- Santiago M, Ramírez-Sarmiento CA, Zamora RA & Parra LP (2016) Discovery, molecular mechanisms, and industrial applications of cold-active enzymes. *Front Microbiol* **7**, 1408–1440.
- Feller G (2010) Protein stability and enzyme activity at extreme biological temperatures. *J Phys Condens Matter* **22**, 323101.
- Åqvist J, Isaksen GV & Brandsdal BO (2017) Computation of enzyme cold adaptation. *Nat Rev Chem* **1**, 1–14.
- Siddiqui KS & Cavicchioli R (2006) Cold-adapted enzymes. *Annu Rev Biochem* **75**, 403–433.
- Georlette D, Blaise V, Collins T, D’Amico S, Gratia E, Hoyoux A, Marx JC, Sonan G, Feller G & Gerday C (2004) Some like it cold: biocatalysis at low temperatures. *FEMS Microbiol Rev* **28**, 25–42.
- Pischedda A, Ramasamy KP, Mangiagalli M, Chiappori F, Milanesi L, Miceli C, Pucciarelli S & Lotti M (2018) Antarctic marine ciliates under stress: superoxide dismutases from the psychrophilic *Euplotes focardii* are cold-active yet heat tolerant enzymes. *Sci Rep* **8**, 1–13.
- Merlino A, Krauss IR, Castellano I, De Vendittis E, Rossi B, Conte M, Vergara A & Sica F (2010) Structure and flexibility in cold-adapted iron superoxide dismutases: the case of the enzyme isolated from *Pseudoalteromonas haloplanktis*. *J Struct Biol* **172**, 343–352.
- Ding H, Zeng Q, Zhou L, Yu Y & Chen B (2017) Biochemical and structural insights into a novel thermostable β-1, 3-galactosidase from *Marinomonas* sp. BSi20414. *Mar Drugs* **15**, 13.
- Oikawa T, Kazuoka T & Soda K (2003) Paradoxical thermostable enzymes from psychrophile: molecular characterization and potentiality for biotechnological application. *J Mol Catal B Enzym* **23**, 65–70.
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V & Henrissat B (2008) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37**, D233–D238.
- Henrissat B & Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* **7**, 637–644.
- Davies G & Henrissat B (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3**, 853–859.
- Husain Q (2010) β Galactosidases and their potential applications: a review. *Crit Rev Biotechnol* **30**, 41–62.
- Hidaka M, Fushinobu S, Ohtsu N, Motoshima H, Matsuzawa H, Shoun H & Wakagi T (2002) Trimeric crystal structure of the glycoside hydrolase family 42 β-galactosidase from *Thermus thermophilus* A4 and the structure of its complex with galactose. *J Mol Biol* **322**, 79–91.
- Jacobson RH, Zhang XJ, DuBose RF & Matthews BW (1994) Three-dimensional structure of β-galactosidase from *E. coli*. *Nature* **369**, 761–766.
- Rojas AL, Nagem RAP, Neustroev KN, Arand M, Adamska M, Eneyskaya EV, Kulminkaya AA, Garratt RC, Golubev AM & Polikarpov I (2004) Crystal structures of β-galactosidase from *Penicillium* sp. and its complex with galactose. *J Mol Biol* **343**, 1281–1292.
- Maksimainen M, Hakulinen N, Kallio JM, Timoharju T, Turunen O & Rouvinen J (2011) Crystal structures of *Trichoderma reesei* β-galactosidase reveal conformational changes in the active site. *J Struct Biol* **174**, 156–163.
- Barrett T, Suresh CG, Tolley SP, Dodson EJ & Hughes MA (1995) The crystal structure of a cyanogenic β-glucosidase from white clover, a family 1 glycosyl hydrolase. *Structure* **3**, 951–960.
- Henrissat B, Callebaut I, Fabrega S, Lehn P, Mornon JP & Davies G (1995) Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci USA* **92**, 7090–7094.
- Schwab C, Sørensen KI & Gänzle MG (2010) Heterologous expression of glycoside hydrolase family 2 and 42 β-galactosidases of lactic acid bacteria in *Lactococcus lactis*. *Syst Appl Microbiol* **33**, 300–307.
- Maksimainen M, Paavilainen S, Hakulinen N & Rouvinen J (2012) Structural analysis, enzymatic characterization, and catalytic mechanisms of β-galactosidase from *Bacillus circulans* sp. alkalophilus. *FEBS J* **279**, 1788–1798.

- 28 Fan Y, Yi J, Hua X, Feng Y, Yang R & Zhang Y (2016) Structure analysis of a glycosides hydrolase family 42 cold-adapted β -galactosidase from *Rahnella* sp. R3. *RSC Adv* **6**, 37362–37369.
- 29 Viborg AH, Fredslund F, Katayama T, Nielsen SK, Svensson B, Kitaoka M, Lo Leggio L & Abou Hachem M (2014) A β 1-6/ β 1-3 galactosidase from *Bifidobacterium animalis* subsp. lactis B 1-04 gives insight into sub-specificities of β -galactoside catabolism within *Bifidobacterium*. *Mol Microbiol* **94**, 1024–1040.
- 30 Solomon HV, Tabachnikov O, Lansky S, Salama R, Feinberg H, Shoham Y & Shoham G (2015) Structure–function relationships in Gan42B, an intracellular GH42 β -galactosidase from *Geobacillus stearothermophilus*. *Acta Crystallogr D Biol Crystallogr* **71**, 2433–2448.
- 31 Fan Y, Hua X, Zhang Y, Feng Y, Shen Q, Dong J, Zhao W, Zhang W, Jin Z & Yang R (2015) Cloning, expression and structural stability of a cold-adapted β -galactosidase from *Rahnella* sp. R3. *Protein Expr Purif* **115**, 158–164.
- 32 Karan R, Capes MD, DasSarma P & DasSarma S (2013) Cloning, overexpression, purification, and characterization of a polyextremophilic β -galactosidase from the Antarctic haloarchaeon *Halorubrum lacusprofundi*. *BMC Biotechnol* **13**, 3.
- 33 Gutshall KR, Trimbur DE, Kasmir JJ & Brenchley JE (1995) Analysis of a novel gene and beta-galactosidase isozyme from a psychrotrophic *Arthrobacter* isolate. *J Bacteriol* **177**, 1981–1988.
- 34 Hu JM, Li H, Cao LX, Wu PC, Zhang CT, Sang SL, Zhang XY, Chen MJ, Lu JQ & Liu YH (2007) Molecular cloning and characterization of the gene encoding cold-active β -galactosidase from a psychrotrophic and halotolerant *Planococcus* sp. L4. *J Agric Food Chem* **55**, 2217–2224.
- 35 Sheridan PP & Brenchley JE (2000) Characterization of a salt-tolerant family 42 β -galactosidase from a psychrophilic Antarctic *Planococcus* isolate. *Appl Environ Microbiol* **66**, 2438–2444.
- 36 Hildebrandt P, Wanarska M & Kur J (2009) A new cold-adapted β -D-galactosidase from the Antarctic *Arthrobacter* sp. 32c—gene cloning, overexpression, purification and properties. *BMC Microbiol* **9**, 151.
- 37 Ohtsu N, Motoshima H, Goto K, Tsukasaki F & Matsuzawa H (1998) Thermostable β -galactosidase from an extreme thermophile, *Thermus* sp. A4: enzyme purification and characterization, and gene cloning and sequencing. *Biosci Biotechnol Biochem* **62**, 1539–1545.
- 38 Tabachnikov O & Shoham Y (2013) Functional characterization of the galactan utilization system of *Geobacillus stearothermophilus*. *FEBS J* **280**, 950–964.
- 39 Pucciarelli S, Devaraj RR, Mancini A, Ballarini P, Castelli M & Schrallhammer M, Petroni G & Miceli C (2015) Microbial consortium associated with the Antarctic marine ciliate *Euplotes focardii*: an investigation from genomic sequences. *Microb Ecol* **70**, 484–497.
- 40 Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, Brown SD, Chang H-Y, El-Gebali S & Fraser MI (2018) InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res* **44**, D279–D285.
- 41 Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M & Sangrador-Vegas A (2016) The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279–D285.
- 42 Valbonesi A & Luporini P (1993) Biology of *Euplotes focardii*, an Antarctic ciliate. *Polar Biol* **13**, 489–493.
- 43 Lonhienne T, Gerday C & Feller G (2000) Psychrophilic enzymes: revisiting the thermodynamic parameters of activation may explain local flexibility. *Biochim Biophys Acta* **1543**, 1–10.
- 44 Teplitsky A, Mechaly A, Stojanoff V, Sainz G, Golan G, Feinberg H, Gilboa R, Reiland V, Zolotnitsky G & Shallom D (2004) Structure determination of the extracellular xylanase from *Geobacillus stearothermophilus* by selenomethionyl MAD phasing. *Acta Crystallogr D Biol Crystallogr* **60**, 836–848.
- 45 Holm L & Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* **38**, W545–W549.
- 46 Krissinel E & Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* **372**, 774–797.
- 47 Nygaard M, Kragelund BB, Papaleo E & Lindorff-Larsen K (2017) An efficient method for estimating the hydrodynamic radius of disordered protein conformations. *Biophys J* **113**, 550–557.
- 48 Alquati C, De Gioia L, Santarossa G, Alberghina L, Fantucci P & Lotti M (2002) The cold-active lipase of *Pseudomonas fragi*: Heterologous expression, biochemical characterization and molecular modeling. *Eur J Biochem* **269**, 3321–3328.
- 49 Brunialti EAS, Gatti-Lafranconi P & Lotti M (2011) Promiscuity, stability and cold adaptation of a newly isolated acylaminoacyl peptidase. *Biochimie* **93**, 1543–1554.
- 50 Ness JE, Welch M, Giver L, Bueno M, Cherry JR, Borchert TV, Stemmer WPC & Minshull J (1999) DNA shuffling of subgenomic sequences of subtilisin. *Nat Biotechnol* **17**, 893–896.
- 51 Miyazaki K, Wintrode PL, Grayling RA, Rubingh DN & Arnold FH (2000) Directed evolution study of temperature adaptation in a psychrophilic enzyme. *J Mol Biol* **297**, 1015–1026.
- 52 Brocca S, Ferrari C, Barbiroli A, Pesce A, Lotti M & Nardini M (2016) A bacterial acyl aminoacyl peptidase

- couples flexibility and stability as a result of cold adaptation. *FEBS J* **283**, 4310–4324.
- 53 Zanhörlein LM, De Giuseppe PO, Honorato RV, Tonoli CCC, Fattori J, Crespin E, De Oliveira PSL, Ruller R & Murakami MT (2016) Oligomerization as a strategy for cold adaptation: structure and dynamics of the GH1 β -glucosidase from *Exiguobacterium antarcticum* B7. *Sci Rep* **6**, 1–14.
- 54 Skalova T, Dohnalek J, Spiwok V, Lipovova P, Vondráčková E, Petrokova H, Dušková J, Strnad H, Kralova B & Hašek J (2005) Cold-active β -galactosidase from *Arthrobacter* sp. C2–2 forms compact 660 kDa hexamers: crystal structure at 1.9 Å resolution. *J Mol Biol* **353**, 282–294.
- 55 Tronelli D, Maugini E, Bossa F & Pascarella S (2007) Structural adaptation to low temperatures—analysis of the subunit interface of oligomeric psychrophilic enzymes. *FEBS J* **274**, 4595–4608.
- 56 Bernhardsgrütter I, Vögeli B, Wagner T, Peter DM, Cortina NS, Kahnt J, Bange G, Engilberge S, Girard E & Riobé F (2018) The multicatalytic compartment of propionyl-CoA synthase sequesters a toxic metabolite. *Nat Chem Biol* **14**, 1127–1132.
- 57 Jung T & Grune T (2013) The proteasome and the degradation of oxidized proteins: part I—structure of proteasomes. *Redox Biol* **1**, 178–182.
- 58 Paredes DI, Watters K, Pitman DJ, Byströff C & Dordick JS (2011) Comparative void-volume analysis of psychrophilic and mesophilic enzymes: structural bioinformatics of psychrophilic enzymes reveals sources of core flexibility. *BMC Struct Biol* **11**, 42.
- 59 Fields PA, Dong Y, Meng X & Somero GN (2015) Adaptations of protein structure and function to temperature: there is more than one way to ‘skin a cat’. *J Exp Biol* **218**, 1801–1811.
- 60 Struvay C & Feller G (2012) Optimization to low temperature activity in psychrophilic enzymes. *Int J Mol Sci* **13**, 11643–11665.
- 61 Feller G & Gerday C (2003) Psychrophilic enzymes: hot topics in cold adaptation. *Nat Rev Microbiol* **1**, 200–208.
- 62 Yuan T, Yang P, Wang Y, Meng K, Luo H, Zhang W, Wu N, Fan Y & Yao B (2008) Heterologous expression of a gene encoding a thermostable β -galactosidase from *Alicyclobacillus acidocaldarius*. *Biotechnol Lett* **30**, 343–348.
- 63 Godoy AS, Camilo CM, Kadowaki MA, Muniz HS, Espírito Santo M, Murakami MT, Nascimento AS & Polikarpov I (2016) Crystal structure of β 1 \rightarrow 6-galactosidase from *Bifidobacterium bifidum* S17: trimeric architecture, molecular determinants of the enzymatic activity and its inhibition by α -galactose. *FEBS J* **283**, 4097–4112.
- 64 Juajun O, Nguyen T-H, Maischberger T, Iqbal S, Haltrich D & Yamabhai M (2011) Cloning, purification, and characterization of β -galactosidase from *Bacillus licheniformis* DSM 13. *Appl Microbiol Biotechnol* **89**, 645–654.
- 65 Seemann T (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069.
- 66 Studier FW (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr Purif* **41**, 207–234.
- 67 Eftink MR & Ghiron CA (1976) Exposure of tryptophanyl residues in proteins. Quantitative determination by fluorescence quenching studies. *Biochemistry* **15**, 672–680.
- 68 Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125–132.
- 69 Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW & McCoy A (2011) Overview of the CCP4 suite and current developments. *Acta Crystallographica Section D* **67**, 235–242.
- 70 McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC & Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**, 658–674.
- 71 Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132.
- 72 Murshudov GN, Vagin AA & Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240–255.
- 73 Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS & Richardson DC (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12–21.
- 74 Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B & Lindahl E (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25.
- 75 Best RB, Zheng W & Mittal J (2014) Balanced protein–water interactions improve properties of disordered proteins and non-specific protein association. *J Chem Theory Comput* **10**, 5113–5124.
- 76 Abascal JLF & Vega C (2005) A general purpose model for the condensed phases of water: TIP4P/2005. *J Chem Phys* **123**, 234505.
- 77 Hopkins CW, Le Grand S, Walker RC & Roitberg AE (2015) Long-time-step molecular dynamics through hydrogen mass repartitioning. *J Chem Theory Comput* **11**, 1864–1874.
- 78 Suchard MA & Redelings BD (2006) BAli-Phy: simultaneous Bayesian inference of alignment and phylogeny. *Bioinformatics* **22**, 2047–2048.

- 79 Le SQ & Gascuel O (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25**, 1307–1320.
- 80 Bansal MS, Kellis M, Kordi M & Kundu S (2018) RANGER-DTL 2.0: rigorous reconstruction of gene-family evolution by duplication, transfer, and loss. *Bioinformatics* **34**, 3214–3216.
- 81 Lartillot N, Lepage T & Blanquart S (2009) PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286–2288.
- 82 Lartillot N & Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21**, 1095–1109.
- 83 Thorne JL, Kishino H & Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* **15**, 1647–1657.
- 84 Gasteiger E, Hoogland C, Gattiker A, Wilkins MR, Appel RD & Bairoch A (2005) Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook*, pp. 571–607. Humana Press, Totowa, NJ: Springer.
- 85 Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD & Wishart DS (2003) VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res* **31**, 3316–3319.
- 86 Piovesan D, Minervini G & Tosatto SCE (2016) The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res* **44**, W367–W374.
- 87 DeLano WL (2002) The PyMOL molecular graphics system. <http://www.pymol.org>
- 88 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC & Ferrin TE (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.